(spatial image) and the video are always consistent. This works as long as there are gaps **72**, **74** in between the areas that the cameras **30**, **32** can see, and when moving audio source directions fall to these gaps; the direction can be changed. Of course, the audio source direction can also be changed when the source is visible in a camera, but that causes an inconsistency between the spatial image and the video. In a typical scenario, the directions of ambient audio sources would be modified; the speaker typically moves less and remains in the view of one camera only.

[0053] Features as described herein may use capturing 5.1 signals using only 3 microphones in a mobile device such as described in U.S. patent publication No. US 2013/0044884 A1 which is hereby incorporated by reference in its entirety. It is also possible to use different panning functions for different outputs. In this embodiment (referring to FIG. **8**) stereo audio is used. This embodiment (referring to FIG. **8**) would not work with 5.1 audio because the playback direction of audio objects would need to switch from back to front when the camera is switched (from front facing camera to main camera or vice versa). The signals captured by the 3 microphones are firstly converted into Mid and Side signals together with side information alpha. Alpha α_b describes the dominant sound direction for each frequency band b. In order to create an audio signal, where the sound image remains static even when the user switches between the front and the rear camera, alpha may be modified in the following way (band index b is omitted for the sake of simplicity):

$$\hat{\alpha} = \begin{cases} \alpha, & A < \alpha \leq 180^\circ - A \\ \emptyset, & 180^\circ - A < \alpha \leq 180^\circ + A \\ 180^\circ - \alpha, & 180^\circ + A < \alpha \leq -A \\ \emptyset, & -A < \alpha \leq A \end{cases} \quad (1)$$

This results in the audio source directions, that fall into the areas **72**, **74** in FIG. **9**, being modified.

[0054] Objects directly to the left and right may be played back from both left and right; therefore they are given direction ø that is the notation for unknown direction. The area for the unknown direction sector is $-A<\alpha<<A$ and $180°-A<\alpha<<180°+A$. A is typically 45°, however, other values for A may also be used, such as when different panning laws are used for example. One example of a panning law that can be used, in particular if A=45°, is the well known sine-cosine panning law. Mid channel panned to Left and Right output channels is then:

$$L^b = \cos(\hat{\alpha}_b) M^b$$

$$R^b = \cos(\hat{\alpha}_b) M^b \quad (2)$$

After this the decorrelated Side signal is added to both Left and Right channel, which are transmitted and played back.

[0055] With the example described above, audio objects directly to the left or right of the device need to be played back from both left and right; otherwise the playback direction of those audio objects will switch places when the camera is switched. This is not a big problem, since those objects are not visible in either camera. Referring also to FIG. **11**, alternatively it is possible to focus on making the spatial audio image stationary on one side of the camera (in front of or behind) while letting the playback direction of audio objects on the three other sides switch from left to right and vice versa when the camera is switched. This can be achieved by compressing the audio object playback locations to one point on one side of the camera while keeping the locations "as is" on the other sides.

[0056] In practice, this would usually be done so that on the front facing camera side all audio objects that are visible in the front facing camera would have their audio always played back from the center. Usually, there is only the person on the front facing camera side, thus, compressing the direction of his voice to the center is natural. Audio objects on other sides of the device would then be played back from the direction that corresponds to the directions seen on the rear camera. This can be done by replacing Formula 1 with modifying a with the function illustrated in FIG. **10**; while keeping other processing as in FIG. **8**. FIG. **10** is a curve depicting the modification of a.

[0057] Referring also to FIG. **12**, an alternative example is to rotate the auditory spatial image slowly when camera use is switched between or among cameras. For example, a user first takes a video with a first camera, and the audio spatial image is consistent with the first camera. The user then switches to a second camera. After the switch the audio spatial image is (slowly) rotated until it becomes consistent with the second camera.

[0058] The above described examples have focused on devices that have two cameras. However, features as described herein may easily be extended to an apparatus having more than two cameras. The cameras do not need to be in a single plane. The same principles as presented here may be used for cameras that are not on a single plane. Alternative embodiments here are not limited to mobile phones. Any device with two or more cameras can be used. It is possible to add means for audio object separation or directional microphones to separate the speech and ambient objects. For example, similar to the embodiment described above with regard to FIG. **6**, features could be implemented in the following two ways with regard to FIGS. **13** and **14**. FIG. **13** shows using audio object separation. Audio is separated into objects from the front and rear cameras as indicated by block **76**, objects from the side of the rear camera **30** may be attenuated as indicated at block **78** depending upon which camera is being used, and signals can be combined **80** for an output audio **82**. FIG. **14** shows using directional microphones. Signals **44**, **45** from the microphones directed to the left and right of the rear camera may be attenuated as indicated by block **84** based upon which camera is being used, and signals **85** from a microphone directed to the same direction as the front camera **32** may be combined **86**, **87** with the signals output from **84** to generate the output left and right channels **46**, **47**.

[0059] Audio object separation can also be used to implement embodiments similar to that described above with respect to FIGS. **8-11**. A method may be used to convert a multi-microphone signal into audio objects and their trajectories. The trajectories are the time dependent directions of each object. The direction is typically indicated as an angle (or two angles, azimuth and elevation, in case of full 3D trajectories) with respect to the device. The angle of each object can then be modified using Formula 1 or FIG. **10**. See FIG. **15** for an example block diagram of this implementation. Signals from the microphones can be separated into objects and their time dependent directions as indicated by block **88**, such as use of features as described in International patent publication No. WO 2014/147442 A1 for example, which is hereby incorporated by reference in its entirety. The direction